

University of St. Thomas, Minnesota UST Research Online

Operations and Supply Chain Management Faculty
Publications

Operations and Supply Chain Management

2012

A Model for Efficiency-Based Resource Integration in Services

Sheneeta W. White

University of St. Thomas, Minnesota, whit6237@stthomas.edu

Ralph D. Badinelli

Virginia Tech, ralphb@vt.edu

Follow this and additional works at: <http://ir.stthomas.edu/ocbopmtpub>



Part of the [Business Administration, Management, and Operations Commons](#), and the
[Management Sciences and Quantitative Methods Commons](#)

Recommended Citation

White, Sheneeta W. and Badinelli, Ralph D., "A Model for Efficiency-Based Resource Integration in Services" (2012). *Operations and Supply Chain Management Faculty Publications*. 17.
<http://ir.stthomas.edu/ocbopmtpub/17>

This Article is brought to you for free and open access by the Operations and Supply Chain Management at UST Research Online. It has been accepted for inclusion in Operations and Supply Chain Management Faculty Publications by an authorized administrator of UST Research Online. For more information, please contact libroadmin@stthomas.edu.



Innovative Applications of O.R.

A model for efficiency-based resource integration in services

Sheneeta W. White^{a,*}, Ralph D. Badinelli^{b,1}^a University of St. Thomas, Operations and Supply Chain Management Department, 2115 Summit Ave, St. Paul, MN 55105, United States^b Virginia Tech, Department of Business Information Technology (0235), 1007 Pamplin Hall, Blacksburg, VA 24061, United States

ARTICLE INFO

Article history:

Received 25 October 2010

Accepted 6 September 2011

Available online 16 September 2011

Keywords:

OR in manpower planning

Services

Coproduction

Dynamic programming

ABSTRACT

Service processes, such as consulting, require coordinated efforts from the service recipient (client) and the service provider in order to deliver the desired output – a process known as resource integration. Client involvement directly affects the efficiency of service processes, thereby affecting capacity decisions. We present a mathematical model of the resource-integration decision for a service process through which the client and the service provider co-produce resource outputs. This workforce planning model is unique because we include the extent of client involvement as a policy variable and introduce to the resource-planning model efficiency and quality performance measures, which are functions of client involvement. The optimization of resource planning for services produces interesting policy prescriptions due to the presence of a client-modulated efficiency function in the capacity constraint and subjective client value placed on participation in the service process. The primary results of this research are optimal decision rules that provide insights into the optimal levels of client involvement and provider commitment in resource integration.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Inherent in complex, high-contact client services, such as consulting and business services, is a close client-provider relationship. This relationship is a determinant of the successful delivery of the service. The service system for a consulting job, for example, typically starts with a contract or service-level agreement between the client and the service provider, which describes the acceptable lead time of the job, the structure of the job and payment, and the responsibilities of all parties involved (Dietrich, 2006). Client resources are needed to provide information about the technical and business needs of the client's organization and to address any problems that may arise during the service process. Throughout this service system the client is an active participant. Our model captures the involvement of the client when the client is seen as a co-producer of the service-system outputs. We assume that, for each client, there is an established agreement regarding deliverables and a delivery date for the work, and that the service processes that are needed to complete the job have been identified. However, there has been no agreement regarding the level of effort of the client.

In this paper we develop a resource-integration model for complex, high-contact client services, and derive general forms of

policies for client involvement and provider involvement in a service process. We shall refer to this model as the co-production resource integration model (CORIM). CORIM is a deterministic, resource-planning model. We explore CORIM both theoretically and experimentally to show the effects on policy of varying the labor constraints, process efficiency and process quality. We use dynamic programming to solve the resource integration problem. The contributions of this paper include:

- A resource-integration model that describes the co-production of outputs by integration of resource inputs from the service provider and the service recipient.
- The inclusion in the model of client involvement as a policy variable and client value as a performance measure, which makes the resource-integration plan the result of a collaborative decision that pursues the interests of both the service provider and the service recipient.
- The introduction of client-modulated efficiency and quality performance measures in the resource-integration model.
- The exposition of tradeoffs and policy forms which are unique to service processes due to the need to coordinate resource inputs from the service provider and the service recipient.
- Derivation of an optimal resource-integration policy and proof of a stationary policy form for the multi-period, resource-integration problem.

Business-to-business services and other client-intensive services are managed across multiple deliverables for multiple clients

* Corresponding author. Tel.: +1 (651) 962 5438; fax: +1 651 962 5093.

E-mail addresses: whit6237@stthomas.edu (S.W. White), ralphb@vt.edu (R.D. Badinelli).

¹ Tel.: +1 (540) 231 7688; fax: +1 (540) 231 3752.

over multiple time periods. The ultimate goal of this research stream is the creation of model support for resource-integration planning in such industries. The current paper lays a foundation stone of this model development by creating and analyzing a model for a single process stage of a service system. A sequel will extend the results presented herein to the multi-stage case.

We review relevant literature in Section 2. In Section 3, we present our model for resource integration and derive basic characteristics of the feasible region. In Section 4, we derive the form of the optimal policy through a decomposition method. In Section 5, we present managerial interpretations of the optimal resource-integration policies. Section 6 discusses the contribution of the research and future research directions.

2. Related literature

The current study extends literature on resource planning of service systems, resource flexibility, and efficiency-based resource planning.

Resource planning is a sequential decision process that strives to apply an organization's capacity most efficiently to meet demand. The seminal work of Holt, Modigliani, Muth, and Simon (HMMS) (Holt et al., 1955, 1956) laid the foundation for planning and control models. Gaimon and Thompson (1984), Anderson (2001), Anderson et al. (2006), and Bordoloi and Matsuo (2001) extend the HMMS model to include multi-sourced employee capacity in resource-planning models for services. Although the aforementioned models capture varying resource capabilities, we treat the variations in client capabilities more robustly by including in our model efficiency and quality measures which are functions of client involvement in the service system. Through these functions we can explicitly capture the nature of co-production of output resources and co-creation of value. Martin et al. (2001) argues that a measurement of productivity that does not capture the client side of the service encounter is inadequate when talking about a business service such as consulting.

A natural extension to planning models with resources from multiple sources is allowance of resource flexibility. With flexible resources, processes can become more efficient. A few well-known papers on workforce agility in a service process are Abernathy et al. (1973), Campbell (1999), Brusco and Johns (1998), Van Oyen et al. (2001), Hopp and Van Oyen (2004), and Hopp et al. (2004). Although these papers show that flexible resources can be used to accomplish differing tasks, they fail to capture the shared value component of coproduced services. Roels et al. (2010) and Xue and Field (2008) capture service coproduction in which the service recipients make choices about the engagement of various service providers as the recipient's self interests dictate. By contrast, the context for the model in the current paper is that of a contracted engagement between the service provider and the client such as those in the service industries of consulting, IT development and other knowledge-based, co-creation services. Generally, once the client commits to the service, the service provider is responsible for managing the service project, which includes planning the extent and scope of client involvement. In making this project plan, the provider specifies the resource integration that will lead to successful service delivery.

Data Envelopment Analysis (DEA), a methodology that was designed for measuring the relative efficiency of decision making units without assumptions are made about the underlying transformation process has been used for efficiency-based resource allocation. Athanassopoulos (1995) and Beasley (2003) model a form of a transformation function by using targets on inputs and outputs while Golany and Tamir (1995) and Korhonen and Syrjänen (2004) use proportional bounds and scaling. In Athanassopoulos (1998)

and Golany et al. (2006), however, an explicit linear transformation function is defined. Although DEA can be used successfully for capturing the effects multiple inputs and multiple outputs on planning and control decisions, it falls short in capturing the nonlinearities that exist in service firms and the need for information regarding the transformation function. The model presented in this paper captures the nonlinearity of service processes through the client-modulated efficiency and quality functions.

To our knowledge, our work is one of the first initiatives to create a mathematical model of client involvement in a service system using explicit efficiency and quality functions. The efficiency and quality functions in this paper are constructed along the lines of the production-function attributes used in Gaimon (1997), Carrillo and Gaimon (2004), and Napoleon and Gaimon (2004). Unlike these workforce planning models for services, our efficiency function modulates a capacity constraint in a planning model, which also includes demand, inventory and backlog of a service. These features produce a model that is more realistic and relevant to the management of business services.

3. Co-production resource integration model

Key components of the model are measures of efficiency and quality of the service process. These measures are functions of the level of client skill and of client intensity. We define client intensity as the ratio of the client's labor input to the service system to the provider's labor input to the service system. Literature reveals additional insights into the effects of client intensity on efficiency and quality. Chase (1978) determined that at some level, further client involvement is either ineffective or detrimental. Therefore, there exists a level of client intensity at which efficiency reaches, perhaps asymptotically, its maximum or saturation level. Consequently, the efficiency and quality functions of our model each strictly increase to a maximum value of 1.0 as a function of client intensity. Gaimon (1997) and Napoleon and Gaimon (2004) posit a set of general, reasonable assumptions about the functional dependencies of efficiency and quality. We augment this structure through the introduction of leverage that is provided by client participation in the service process. Below, we summarize the assumed characteristics of these functions.

- Efficiency and quality are strictly increasing and concave in client intensity.
- Efficiency and quality are bounded above by a saturation level.
- Efficiency and quality are strictly increasing and concave in provider skill and in client skill.
- Efficiency and quality are strictly increasing in the quality of the other resource inputs.

These assumptions determine characteristics of the shape of the efficiency and quality functions. See Fig. 1 for an example. In the current paper we are concerned with the dependency of the efficiency and quality functions on client intensity.

We define the deliverables of a service as service components. In the case of a software consulting firm, for example, service components could be database designs, web-page construction, code writing and testing, etc. We assume that the service provider delivers a service through a service system consisting of a network of linked processes – each process delivering one type of service component. Each process of the network requires a certain number of “cycles” per unit of the service component that is delivered to the client. The cycle of each process has a standard labor requirement.

A key element of the provider's resource integration decision is the provider's assessment of the client's costs of involvement and

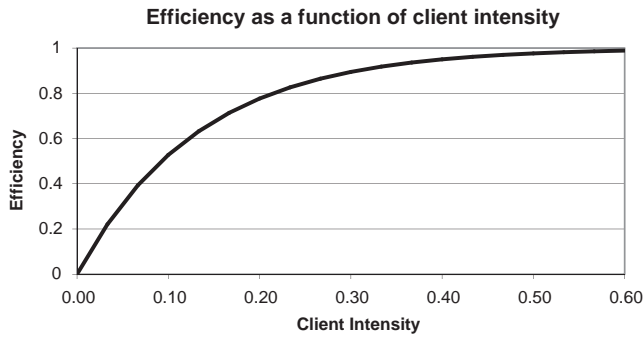


Fig. 1. Example of an efficiency function.

the value to the client of involvement. Certainly, the time commitment of the client must be recognized as a marginally increasing penalty, as this commitment increasingly interferes with other client resource needs. However, it is well known that a client's involvement in service processes yields benefits in the form of stronger acceptance and ownership of the service deliverables, willingness to embrace changes in behavior that the service deliverables require and perceived value of the service. The modeling of the client's value function is complex and the subject of much current research, placing this modeling beyond the scope of the current paper (Heitz et al. (2009), Tang (2009) and Strong et al. (2009)). We only impose the modest assumption that the client's value of involvement in the production of service is marginally decreasing.

These penalties and benefits taken together can be viewed as a cost function on the client's involvement in the service system. The provider must assess this cost function in order to plan resource integration – a challenging task that is familiar to every consulting firm and software development team. The development of assessment models is beyond the scope of the current paper. However, the relevance of this assessment to resource planning and the impact of assessment accuracy are demonstrated by the model that is presented herein.

In consideration of the common practice of assigning service workers to multiple projects, implying part-time assignment to any given stage of a given service, we measure the provider's workforce in terms of the continuous measure of full-time equivalents (FTE) instead of the discrete measure of number of personnel. Similarly, labor-assignment variables are measured in terms of FTE.

The model development presented below is that of the resource-integration plan for a single service process. We establish a stationary form for the optimal policy of the multi-period, resource-integration problem through the modeling of a given time period as a stage of a dynamic program. All of the decision variables, performance measures and parameters are defined for the current time period for this process. In order to simplify notation we suppress the subscripts that would explicitly identify the time period and process.

3.1. Definitions and notation

Decision Variables

g	generation of process cycles of the service process (# of completed cycles)
w	number of provider FTEs assigned to the service process
h	number of provider FTEs hired for the service process
f	number of provider FTEs fired from the service process
y	client intensity = the ratio of client time to provider time spent on the service process

We use the terms “hire” and “fire” to represent the assignment of personnel to a service process and the removal of personnel from a process, respectively. This assignment and removal can take many forms across different contexts of service provision. In some cases, these actions can literally be in the form of hiring and firing or laying-off. In other contexts, the “hiring” is in the form of assigning a portion of a service-worker's time to a service process and “firing” is in the form of re-assigning this worker. Many different words can be used to indicate these labor decisions, each one relevant to a certain context. We settle the matter by choosing the terms hire and fire in a generic sense.

State Variables

- | | |
|-----|---|
| i | “inventory” of the service component at the end of the current period = the number of units of the service component that are completed earlier than needed |
| b | backlog of the service component at the end of the current period = the number of units of the service component that are overdue |

Note that inventory and backlog are real-valued to reflect the possibility of partial completion of a service component.

Performance Measures

$e(y)$	efficiency of the service process as a function of client intensity
$q(y)$	quality of the service process as a function client intensity
$c_c(yg)$	“cost” of client involvement in the service process = net effect of penalties and benefits of the time spent by the client, yg , in the service process
$c_b(b)$	the loss of value of the service process due to delayed deliverables
$V(n, w)$	present value of the optimal plan after the current period if the current period ends with net inventory, $n = i - b$, and workforce, w
z	total profit of the resource integration plan for the service process over the current and future periods

As we stated earlier, we assume that the efficiency and quality measures are strictly increasing, concave functions of y and that c_c is convex.

The backlog variable measures the number of units of the service component that are not completed by their due dates. We impose a cost on this backlog, c_b , in order to capture the marginally increasing losses of client perceived value as well as postponed revenue to the service provider. Hence, we assume that c_b is convex and increasing. Backlog also increases future labor costs, which are captured by the function, V .

There is no explicit inventory holding cost. If service deliverables are completed prior to or later than the times that they are needed, then the cost of generating these deliverables is recognized in the time periods of generation. Hence, early or tardy generation will affect net present value. We include in the objective function the optimal discounted future profit to capture the effects of the decision in the current period on future optimal performance.

Once the costs of generation in the current period are accounted for, higher inventory or lower backlog at the end of the period has a positive marginal net present value, as it can reduce future labor cost and, potentially, future backlog costs. There is a limit to the

benefit of inventory when the inventory level reaches that which covers all future demand. Define,

n_{\max} = the inventory level that covers all future demand,
 $g_{\max} = r(n_{\max} + d - i_0 + b_0)$
 = the generation that yields a net inventory level of.

From these considerations, we assume that V is differentiable and non-decreasing in $n < n_{\max}$. We assume the V is at least quasi-concave in w , imposing a reasonable and weak condition on the tradeoffs for the existence of an optimal work-force.

Parameters	
c_h	cost of hiring an FTE of the service provider for the service process
c_f	cost of firing an FTE of the service provider from the service process
c_w	cost of wages per FTE-period of the service provider
v	revenue per unit of the service component
d	demand for the service component (units of the service component)
r	required number of cycles of the service process per unit of the service component
r^h	number of standard labor hours required per cycle of the service process
a^w	available provider capacity (labor hours/FTE-period)
a^c	available client capacity for the service (labor hours/period)
\bar{w}	maximum provider workforce level (FTEs)
\bar{b}	maximum allowed backlog of the service process (# units of the service component)
\underline{y}	minimum required client intensity
\bar{y}	maximum allowable client intensity
\underline{q}	minimum required quality level

Note: $r^h g y$ = number of hours allocated to the service by the client.

The resource integration problem can be stated as follows:

Problem P.

$$\begin{aligned}
 &\text{Maximize}_{w,h,f,i,b,y,g} \quad z(w,h,f,i,b,y,g) = dv - c_h h - c_f f \\
 &\quad - c_w w - c_c(yg) - c_b(b) + V(i - b, w), \\
 &\text{Subject to:} \quad w - w_0 - h + f = 0, \quad (1) \\
 &\quad \bar{w} - w \geq 0, \quad (2) \\
 &\quad \bar{b} - b \geq 0, \quad (3) \\
 &\quad a^w w e(y) - r^h g \geq 0, \quad (4) \\
 &\quad a^c - r^h y g \geq 0, \quad (5) \\
 &\quad y - \underline{y} \geq 0, \quad (6) \\
 &\quad \bar{y} - y \geq 0, \quad (7) \\
 &\quad q(y) - \underline{q} \geq 0, \quad (8) \\
 &\quad i_0 - b_0 + g/r - d - i + b = 0, \quad (9) \\
 &\quad g_{\max} - g \geq 0, \quad (10) \\
 &\quad w, h, f, g, y, i, b \geq 0.
 \end{aligned}$$

The objective function maximizes the net present value of all current and future resource commitments, inventory, backlog and revenue. In the spirit of co-production we view this optimization as a joint venture between the provider and the client. Consequently, the client's costs and the provider's costs are combined in the objective function, and the functions, c_c and c_b , reflect client values.

Constraint (1) is a workforce balance constraint. Constraint (2) ensures that the maximum workforce level that the provider can accommodate is not exceeded. Constraint (3) ensures that the number of service jobs tardy (backordered) does not exceed the maximum level that is allowed by the service level agreement.

Constraints (4) and (5) are the provider and client capacity constraints, respectively. The efficiency term in the capacity constraint (4) is needed to represent the effect of client involvement on the effective capacity of the workforce. Note that the boundaries of each of Constraints (4) and (5) define generation as a function of client intensity. We denote these two functions as $g_4(y)$ and $g_5(y)$, respectively, and the inverses of these functions as $y_4(g)$ and $y_5(g)$, respectively.

In Constraints (6) and (7) we have set minimum and maximum amounts of client involvement, respectively. We assume that in business services such as consulting and IT development the client will always be part the service process to a certain extent, and that the client cannot be the sole labor source in any process.

Typically, organizations benchmark themselves against competitors in terms of quality and establish internal quality standards. Constraint (8) imposes a minimum process quality level that must be achieved.

Constraint (9) is the conventional balance equation that establishes the functional dependencies among inventory, backlog, demand and generation. Constraint (10) sets an upper bound on generation that is based on n_{\max} .

Problem P can be solved efficiently with standard non-linear programming methods. However, our purpose in this paper is to reveal the form of the optimal policy. Consequently, we proceed with a detailed analysis of Problem P and its optimization.

3.2. Conditions

In order to ensure non-trivial solutions and to establish tighter bounds on feasible solutions, we derive several conditions on the parameters of the problem.

Condition 1. Constraints (3) and (9) set a lower limit on generation. Define,

$$g_{\min} = r(d - i_0 + b_0 - \bar{b}).$$

We assume that demand is high enough to require $g_{\min} > 0$. Constraint (3) can be replaced with,

$$g - g_{\min} \geq 0. \quad (11)$$

Condition 2. Since $q(y)$ is strictly increasing and bounded above by 1, Constraint (8) can be re-written as a lower bound on the client intensity. There is a unique value of y for which $q(y) = \underline{q}$. We shall denote this value, y_q . If $q(0) > \underline{q}$, then we set $y_q = 0$. The highest level of client intensity that is allowed, \bar{y} , must be able to provide the minimal level of quality. Therefore, $y_q < \bar{y}$. Define,

$$y_{\min} = \max(y, y_q).$$

Constraints (6) and (8) can be replaced with the single constraint,

$$y - y_{\min} \geq 0. \quad (12)$$

Condition 3. Constraint (5) must provide enough capacity to achieve the minimal required generation. This requirement imposes an upper bound on the client intensity. Define,

$$y_5(g) = \frac{a^c}{r^h g},$$

$$y_{\max} = \min(\bar{y}, y_5(g_{\min})).$$

$y_5(g)$ = unique value of client intensity that is specified by the boundary of Constraint (5) for any given generation level, $g > 0$. Therefore, $y_{\min} \leq y \leq y_5(g_{\min})$. Constraint (7) can be replaced with the tighter constraint,

$$y_{\max} - y \geq 0. \quad (13)$$

Condition 4. Constraints (4), (11) and (13) place a lower bound on the provider's workforce level, which we denote, \underline{w} . This condition follows from the facts that, $e(y_{\max})$ is the highest level of efficiency obtainable and g_{\min} is the smallest number of cycles that can be generated. Define,

$$\underline{w} = \frac{r^h g_{\min}}{a^w e(y_{\max})},$$

$$w - \underline{w} \geq 0. \quad (14)$$

Condition 5. Constraints (4), (11) and (12) specify a workforce level below which Constraint (4) cannot provide enough capacity for the minimum required generation without requiring more than the minimum level of client intensity. We denote this workforce level, w' .

$$w' = \frac{r^h g_{\min}}{a^w e(y_{\min})}.$$

Condition 6. Constraints (4), (5), (10) and (12) specify a workforce level, which we denote, w'' , above which Constraint (4) is not binding. Constraints (4), (5) and (10) place upper bounds on the generation. Since the bound on generation that is imposed by Constraint (4) is increasing in y and the bound on generation that is imposed by Constraint (5) is decreasing in y , if $g_5(y_{\min}) < g_4(y_{\min})$ or $g_{\max} < g_4(y_{\min})$ then Constraint (4) is redundant for all y , $y_{\min} \leq y \leq y_{\max}$. Define,

$$w'' = \frac{r^h \min(g_{\max}, g_5(y_{\min}))}{a^w e(y_{\min})}.$$

Condition 7. The marginal value of the workforce to the optimal future resource-integration plan, V , is bounded by the costs of hiring and firing because, for any workforce level, w , this state could be achieved from a workforce level of $w - 1$ and hiring one FTE or from a workforce level of $w + 1$ and firing one FTE. Therefore,

$$-c_f \leq \frac{\partial V}{\partial w} \leq c_h.$$

Condition 8. We can eliminate the variables, h and $i - b$, and the equality Constraints (1) and (9) by substituting for these variables.

$$h = w - w_0 + f.$$

The constraint, $h \geq 0$ becomes,

$$w - w_0 + f \geq 0, \quad (15)$$

$$i - b = i_0 - b_0 + g/r - d.$$

We define $c_b(-i_0 + b_0 - g/r + d) = 0$ for $-i_0 + b_0 - g/r + d < 0$

4. Dynamic program

Problem P is expressed as a multi-period dynamic program. We now decompose this problem further by separating the hiring/firing/workforce decision variables from the rest of the decision variables. The current period's problem is then a two-stage dynamic program. Stage 1 is the optimization of the generation and client plan for a given level of the provider's workforce – **Problem P1**, below. Stage 2 is the optimization of the combined workforce and generation-client plans – **Problem P2** below. The state variable that connects the Stage 2 to Stage 1 is the workforce level, w . Applying **Condition 8** to this two-stage formulation we define,

$$z_1(y, g; w, i_0 - b_0) = -c_c(yg) - c_b(-i_0 + b_0 - g/r + d) + V(i_0 - b_0 + g/r - d, w),$$

$$z_2(w, f; w_0, i_0 - b_0) = z_1^*(w, i_0 - b_0) - (c_f + c_h)f - (c_w + c_h)w + c_h w_0 + d v.$$

Problem P1

$$\max_{y, g} \quad z_1(y, g; w, i_0 - b_0)$$

Subject to : (4), (5), (10), (11), (12), (13),

Problem P2

$$\max_{w, f} \quad z_2(w, f; i_0 - b_0, w_0)$$

Subject to : (2), (14), (15) and

$$f \geq 0. \quad (16)$$

4.1. Optimality conditions for Problem P1

First, we derive the solution of **Problem P1** for the optimal production plan, given a workforce, w , and initial net inventory, $i_0 - b_0$. Once this solution is obtained, we can solve **Problem P2** to find the combination of the optimal provider workforce level and its associated generation-client plan.

We now derive some basic expressions that are related to the optimality condition of **Problem P1**. Define,

$$X(w, i_0 - b_0) = \text{feasible region of Problem P1}$$

The gradient of the objective function for **Problem P1** is,

$$\nabla z_1^T = \left(-c'_c(yg)g, -c'_c(yg)y + \frac{1}{r}c'_b(-i_0 + b_0 - g/r + d) + \frac{1}{r} \frac{\partial V}{\partial n} \right). \quad (17)$$

The Lagrangian for **Problem P1** is,

$$L_1 = z_1 + \lambda_4(a^w w e(y) - r^h g) + \lambda_5(a^c - r^h y g) + \lambda_{10}(g_{\max} - g) + \lambda_{11}(g - g_{\min}) + \lambda_{12}(y - y_{\min}) + \lambda_{13}(y_{\max} - y).$$

The necessary Karush–Kuhn–Tucker (KKT) optimality conditions (see **Winston, 1990**) include the following two, first-order conditions,

$$-c'_c(yg)g + \lambda_4 a^w w \frac{de}{dy} - \lambda_5 r^h g + \lambda_{12} - \lambda_{13} = 0, \quad (18)$$

$$-c'_c(yg)y + \frac{1}{r}c'_b(-i_0 + b_0 - g/r + d) + \frac{1}{r} \frac{\partial V}{\partial n} - \lambda_4 r^h - \lambda_5 r^h y - \lambda_{10} + \lambda_{11} = 0. \quad (19)$$

Fig. 2 shows an example of the constraints of Problem P1, which will aid the reader in understanding the conditions and propositions that follow.

The client value function, $-c_c(yg)$, plays a critical role in determining the optimal policy. From conditions (18) and (19) we can see that the optimal solution to Problem P1 depends significantly on the sign of $-c'_c(yg)$. Accordingly, we recognize the importance of the value of the argument of c_c at which the client value function reaches a maximum, which we define the maximum client value point (MCVP). We define the maximum client value frontier (MCVF) as the locus of points in the $y - g$ plane that satisfy the condition, $-c'_c = 0$. The MCVF is a hyperbola that is defined by,

$$yg = \text{MCVP}. \quad (20)$$

We note that condition (20) is independent of all parameters other than the parameters of the client's cost and benefit functions and g_{\max} . For $g > g_{\max}$, the client would not experience any benefit from generation. Therefore, there is an upper bound on MCVP, specifically,

$$\text{MCVP} \leq y_{\max} g_{\max}. \quad (21)$$

We also note that the hyperbola defined by (20) is either coincident with or non-intersecting with the client capacity constraint (5). The example of Fig. 2 shows three fundamentally different cases of the MCVF. These cases lead to different forms of the optimal solution, which are established by the propositions below. We note that the position of the MCVF relative to the feasible region is determined only by the client's utility function and is independent of all of the parameters of the problem. Therefore, a robust optimization of co-creative resource integration necessitates solutions for all possible client utility functions. The propositions below determine the form of the solution for each possible case.

We can now prove several propositions which lead to a specification of the optimal solution to Problem P1. Proposition 1 establishes an important monotonicity property of the objective function, z_1 , along the MCVF and the capacity constraint, (5).

Proposition 1. Along any arc, $yg = \text{constant}$, z_1 is non-decreasing in g and non-increasing in y .

Proof. The proof is in the Supplemental document. \square

In case 1 of the example of Fig. 2, Proposition 1 implies that the optimal solution lies on the left-most feasible point of the Constraint (5) if this constraint is binding.

Corollary 1. If Constraint (5) is binding at optimality, then an optimal solution is located at the point on this boundary of the minimal feasible value of y and the maximum feasible value of g .

If the maximum client value is attained at higher intensities and generations than are feasible, that is, the MCVF lies above the feasible region, then the optimal solution to Problem P1 is a point on the boundary of the feasible region that bounds client intensity and generation from above. Refer to case 2 of Fig. 2. The only constraints that place an upper bound on the client intensity are Constraints (5) and (13). Along Constraint (5), by Proposition 1, the gradient favors higher generation. Along Constraint (13), the gradient also favors higher generation. Therefore, the optimal solution is found on either of these constraints at the highest feasible generation and client capacity is maximally utilized. We formally establish this fact in Proposition 2.

Proposition 2. If, for all feasible solutions, $yg \leq \text{MCVP}$ then $y^* g^* \geq \{g|(y, g) \in X(w, i_0 - b_0)\}$ and Constraints (5) or (13) binding.

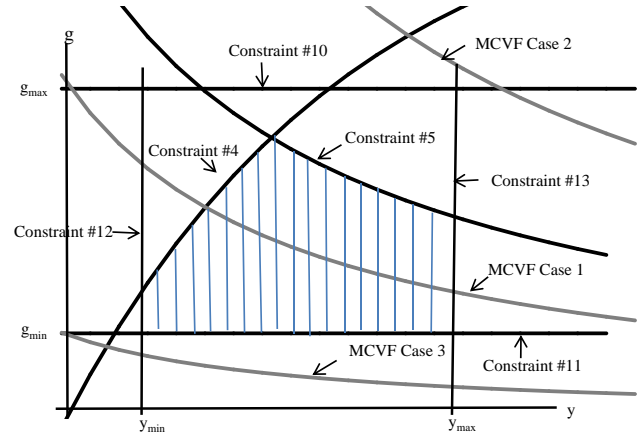


Fig. 2. The feasible region for P1.

Proof. The proof is in the Supplemental document. \square

If the maximum client value is attained at lower intensities and generations than are feasible, that is, the MCVF lies below the feasible region then the optimal solution to Problem P1 is a point on the boundary of the feasible region that provides the minimum feasible client intensity. Refer to case 3 in Fig. 2. In this case, the only constraints that place a lower bound on the client intensity are Constraints (4) and (12) and the provider's capacity is maximally utilized. Proposition 3 establishes the fact that these constraints that are binding in these two cases.

Proposition 3. If, for all feasible solutions, $yg \geq \text{MCVP}$ then $y^* g^* \leq \{y|(y, g) \in X(w, i_0 - b_0) \text{ and Constraints (4) or (12) are binding}\}$.

Proof. The proof is in the Supplemental document. \square

If the MCVF intersects the feasible region, then, by Proposition 1, the optimal solution to Problem P1 is a point on or above this line that provides minimal client intensity and maximum generation. Refer to case 1 of Fig. 2.

Proposition 4. If $yg = \text{MCVP}$ for some $(y, g) \in X(w, i_0 - b_0)$ then $y^* g^* \geq \text{MCVP}$ and Constraints (4), (10) or (12) are binding.

Proof. The proof is in the Supplemental document. \square

4.1.1. optimality conditions for Problem P2

Analysis of the optimality conditions for Problem P2 leads to the proof that the optimal value of the objective function, z_2^* , is non-decreasing in the state variable, $i_0 - b_0$, and quasi-concave in the state variable, w_0 . This fact is instrumental in establishing a stationary policy for the multi-period problem. We begin the analysis with the behavior of the optimal value of the objective function to Problem P1, $z_1^*(w, i_0 - b_0)$, with respect to $i_0 - b_0$.

Proposition 5. $z_1^*(w, i_0 - b_0)$ is non-decreasing in $i_0 - b_0 < n_{\max}$

Proof. The proof is in the Supplemental document. \square

Proposition 5 allows us to determine the behavior of the optimal solution to Problem P2, z_2^* , with respect to $i_0 - b_0$.

Proposition 6. z_2^* is non-decreasing in $i_0 - b_0 < n_{\max}$

Proof. The proof is in the [Supplemental document](#). \square

From the necessary, first-order KKT conditions for [Problem P2](#) we can prove that the optimal value of the objective function of the single-period problem is quasi-concave in the initial work force level. The lagrangian (Winston, 1990) for [Problem P2](#) is,

$$L_2 = z_1^*(w, i_0 - b_0) - (c_w + c_h)w - (c_h + c_f)f + c_h w_0 + \lambda_2(\bar{w} - w) + \lambda_{14}(w - \underline{w}) + \lambda_{15}(w + f - w_0) + \lambda_{16}f, \quad (22)$$

$$\frac{\partial L_2}{\partial w} = 0 \Rightarrow \frac{\partial z_1^*}{\partial w} \Big|_{w=w^*} = c_w + c_h + \lambda_2 - \lambda_{14} - \lambda_{15}, \quad (22)$$

$$\frac{\partial L_2}{\partial f} = 0 \Rightarrow \lambda_{15} + \lambda_{16} = c_h + c_f. \quad (23)$$

Proposition 7. z_2^* is quasi-concave in w_0

Proof. The proof is in the [Supplemental document](#). \square

4.2. Multi-period case

A major result of this paper is the establishment of a stationary policy for the multi-period resource-integration problem. [Proposition 8](#) provides useful bounds on the sensitivity of the optimal value of the objective function of [Problem P2](#) to the initial workforce level.

Proposition 8.

$$-c_f \leq \frac{\partial z_2^*}{\partial w_0} \leq c_h.$$

Proof. The proof is in the [Supplemental document](#). \square

Proposition 9. For every time period, $V(n, w)$ is increasing in n and quasi-concave in w , $-c_f \leq \frac{\partial V}{\partial w} \leq c_h$ for every time period.

Proof. The proof is in the [Supplemental document](#). \square

[Proposition 9](#) implies that [Propositions 1–8](#) hold for every period, the form of the optimal policy as specified by these Propositions is stationary and the policy depends only on the state variables, $w_0, i_0 - b_0$.

5. Managerial interpretations

For any given workforce, a service firm finds itself in one of three cases, identified in each of [Propositions 2–4](#). In setting a policy for client intensity and service generation level, a trade-off between client costs and net inventory costs is made. What distinguishes the three cases is the nature of this trade-off, which we now explain.

5.1. Tradeoffs

Tradeoffs of the resource-integration problem are non-linear and more complex than those of a manufacturing resource plan. We illustrate this phenomenon with an example, using the data shown in [Table 1](#). [Fig. 3](#) illustrates the non-linear tradeoff between client slack capacity and provider slack capacity as functions of client intensity for this example.

The MCVP specifies levels of client intensity and generation at which the trade-off for the client between the cost of involvement and the benefits of involvement are optimal. The tradeoff for the

Table 1

Experimental parameter data.

Hiring cost (c_h)	10,000	Provider capacity (a^w)	160
Firing cost (c_f)	5000	Client capacity (a^c)	160
Cost of wages (c_w)	8000	Max provider workforce (\bar{w})	30
Revenue (v)	20,000	Min client intensity (\underline{y})	20%
Demand (d)	40	Max client intensity (\bar{y})	90%
Required # of cycles (r)	1	Min quality (\underline{q})	70%
# of standard labor hours/cycle (r^h)	80	Max backlog (\bar{b})	10% of demand

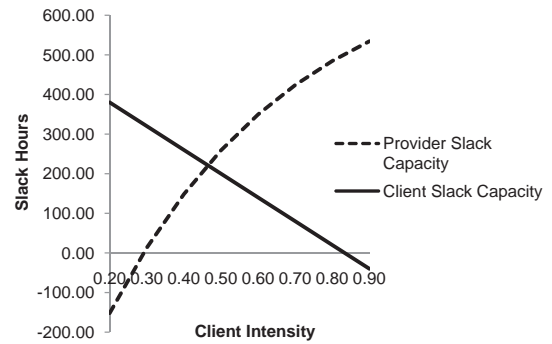


Fig. 3. Tradeoff Illustration.

service provider augments the client's tradeoff with the value of inventory and the cost of the provider workforce. If the optimal tradeoff prescribes a level of client involvement that is above the MCVP, then, in order to support a more overall efficient delivery of the service, the service provider must ask the client to commit more time and resources to the service process than is in the immediate interests of the client.

As educators, consultants and medical-care specialists know, this argument is usually challenging to make. However, the alternatives are sub-optimal policies that consume excessive provider resources. The service provider is well-advised to be aware of the position of the intensity-generation plan relative to the MCVP in order to prepare for the possibility of client dis-satisfaction.

The gradient of z_1 can favor lower client intensity, which implies that quality is sacrificed for cost. This phenomenon is not surprising for a model that places a bound on quality while charging a variable cost for client involvement in the service process. The optimal policy reduces client involvement until either Constraint (4) or Constraint (12) is binding, as these are the only constraints that impose a lower bound on client intensity. Therefore, the resource integrators choose to reduce client involvement in the service process until the minimum allowable quality level is reached or until provider capacity is sufficient to maintain desired generation levels.

An interesting speculation is that the optimal resource integration policy would change in favor of higher client intensity if resource integrators choose to evaluate quality on a ratio scale as opposed to a categorical scale, as a constraint on the quality level (8) implies. In other words, a Taguchi approach to service quality would motivate higher levels of client involvement in the service process. Genichi Taguchi's methodology, widely adopted in manufacturing quality control, views any deviation from quality targets as a financial loss function that is increasing and continuous in the magnitude of the deviation (Evans and Lindsay, 2011).

5.2. Load leveling

If $c_w - c_f \leq \frac{\partial z_1^*}{\partial w} \Big|_{w=w_0} < c_w + c_h$ then, by condition (22), the optimal solution is, $w^* = w_0$, and the provider should maintain the current workforce. This condition is intuitively consistent with all workforce planning policies. That is, the stability of the optimal workforce plan increases with the hiring and firing cost parameters. However, in addition to the conventional motive for not changing the workforce level, diminishing marginal value of client intensity due to the concavity of the efficiency function motivates keeping client intensity down, even at the expense of backlog or lower inventory. It pays to produce less and defer generation to a period when the client's involvement is leveraged better because the generation can take place at a lower level of client intensity. Hence, a load-leveling resource plan is motivated not only by hiring/firing cost, but also by the diminishing effectiveness of client intensity.

5.3. Sensitivity analysis

Sensitivity analysis is essential in the interpretation of the model's results because the model is based on a deterministic representation of the efficiency, quality and client-value functions. Although it is reasonable to assume that other parameters of the model may be estimated with an accuracy that approaches determinism, the same cannot be said of the parameters of these three functions. For most services, the natures of service efficiency and quality are poorly understood and client value is even more of a mystery. Therefore, we should examine the potential effects of errors in estimating these functions.

Errors in estimating the quality function will affect the position of Constraint (12). Very simply, if the quality function is over-estimated, then Constraint (12) will allow levels of client intensity that will not support minimum required levels of quality. If the quality function is under-estimated, then Constraint (12) will require a minimum level of client intensity that exceeds the level that can support the minimum required quality. Considering that most service clients would prefer to err on the side of higher quality, we can advise all users of the model to under-estimate the quality function when there is any doubt about its true form.

The estimation of the efficiency function presents more challenging problems. The effect of the efficiency function on a prescription for a resource-integration policy is seen through the position and shape of Constraint (4). The efficiency function influences the determination of the point on Constraint (4) where an optimal solution is found.

We can deduce certain properties of the sensitivities to mis-estimations in the efficiency function in terms of the comparison of the policy that the mis-estimated model would recommend, (y', g') , to the true optimal policy, (y^*, g^*) . If the efficiency function is over-estimated (under-estimated), then the solution will be set at $y' < y^*, g' > g^*$ ($y' > y^*, g' < g^*$).

In the case of an over-estimated efficiency, the policy prescription attempts to generate more service and engages less client time than is optimal. Of course, the actual service that is generated will be less than expected, and, given the less-than-optimal client intensity, the actual generation will even be less than the optimal generation. The effects of these errors are a lower inventory or a higher backlog level and a higher client expense than are warranted or expected by the resource planner. Higher client costs will be incurred in a future period when the generation shortage is recognized. There could also be a surprising backlog cost or even a violation of the backlog constraint in this case.

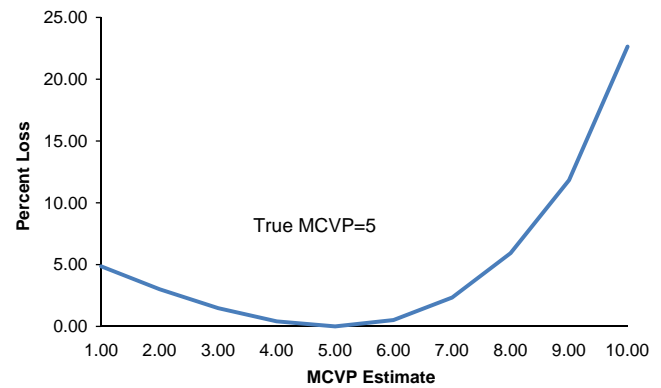


Fig. 4. Sensitivity of optimal objective function to MCVP.

In the case of an under-estimated efficiency, the policy prescription generates less service and engages more client time than is optimal. The effects of these errors are a lower inventory or higher backlog level and a higher client expense than are warranted. However, there are no surprises in terms of differences between actual output and planned output. We conclude that resource integrators should generally err on the side of under-estimating efficiency.

Finally, we recognize the crucial role of estimating the client cost function, c_c , in setting a resource-integration policy. Recall that this function incorporates the client's perceived value of the service, an individualized and context-sensitive function. The many issues involved in this estimation and the myriad potential methodologies for performing the estimation are beyond the scope of this research. However, the model presented herein serves to identify this function and to establish clearly its role in resource-integration planning. Fig. 4 shows an example of the comparison of actual performance of the solution to the resource-integration problem to the optimal performance as a result of mis-estimating the MCVP. The graph shows the percent loss in the objective function due to mis-estimation.

6. Conclusions and future research

In this paper, we have developed a resource-integration model that will serve as a foundation for future service-science research. Our model is unique through its incorporation of the client as a resource in the generation plan. We also capture the effects of client intensity on efficiency and quality in our model. The policy recommendations of the model give service enterprises valuable information regarding their hiring and firing policies, level of client intensity, and service generation. We preserve the notion of inventory in resource planning through the use of discounted cash flows and deferred revenues.

The generality of the results are due to the generally applicable assumptions about the efficiency and quality functions. Therefore, the model is robust with respect to its breadth of applications as well as with respect to the theoretical insights we have derived. A fruitful and exciting direction for future research is the challenge of estimating the client value function, the efficiency function and the quality function. A further extension of this model is the incorporation of client learning and the potential of the service provider to influence changes in the client's value function over time.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ejor.2011.09.009](https://doi.org/10.1016/j.ejor.2011.09.009).

References

- Abernathy, W.J., Baloff, N., et al., 1973. A three-stage manpower planning and scheduling model – A service-sector example. *Operations Research* 21 (3), 693.
- Anderson, E.G., 2001. The nonstationary staff-planning problem with business cycle and learning effects. *Management Science* 47 (6), 817.
- Anderson, E.G., Morrice, D.J., Lundeen, G., 2006. Stochastic optimal control for staffing and backlog policies in a two-stage customized service supply chain. *Production and Operations Management* 15 (2), 262.
- Athanassopoulos, A.D., 1995. Goal programming & data envelopment analysis (GoDEA) for target-based multi-level planning: Allocating central grants to the Greek local authorities. *European Journal of Operational Research* 87 (3), 535–550.
- Athanassopoulos, A.D., 1998. Decision support for target-based resource allocation of public services in multiunit and multilevel systems. *Management Science* 44 (2), 173.
- Beasley, J.E., 2003. Allocating fixed costs and resources via data envelopment analysis. *European Journal of Operational Research* 147 (1), 198–216.
- Bordoloi, S., Matsuo, H., 2001. Human resource planning in knowledge-intensive operations: A model for learning with stochastic turn-over. *European Journal of Operational Research* 130 (1), 169–189.
- Brusco, M.J., Johns, T.R., 1998. Staffing a multiskilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Sciences* 29 (2), 499.
- Campbell, G.M., 1999. Cross-utilization of workers whose capabilities differ. *Management Science* 45 (11), 722–732.
- Carrillo, J.E., Gaimon, C., 2004. Managing knowledge-based resource capabilities under uncertainty. *Management Science* 50 (11).
- Chase, R., 1978. Where does the customer fit in a service operation. *Harvard Business Review*, 137–142.
- Dietrich, B., 2006. Resource planning for business services. *Communication of the ACM* 49 (7), 62–64.
- Evans, J., Lindsay, W.M., 2011. *Managing for quality and performance excellence*, South-Western Cengage Learning, Ohio.
- Gaimon, C., 1997. Planning information technology-knowledge worker systems. *Management Science* 43 (9).
- Gaimon, C., Thompson, G.L., 1984. A distributed parameter cohort personnel planning model that uses cross-sectional data. *Management Science* 30 (6).
- Golany, B., Hackman, S.T., et al., 2006. An efficiency measurement framework for multi-stage production systems. *Annals of Operations Research* 145 (1), 51.
- Golany, B., Tamir, E., 1995. Evaluating efficiency–effectiveness–equality trade-offs: A data envelopment analysis approach. *Management Science* 41 (7), 1172.
- Heitz, C., Dettling, M., Ruckstuhl, A., (2009). Customer lifetime value in contractual settings, paper presented at Annual INFORMS Meeting, San Diego, CA.
- Holt, C.C., Modigliani, F., et al., 1955. A linear decision rule for production and employment scheduling. *Management Science* (pre-1986) 2 (1), 1.
- Holt, C.C., Modigliani, F., et al., 1956. Derivation of a linear decision rule for production and employment. *Management Science* (pre-1986) 2 (2), 159.
- Hopp, W., Tekin, E., Van Oyen, M.P., 2004. Benefits of skill chaining in serial production lines with cross-trained workers. *Management Science* 50 (1), 83–98.
- Hopp, W., Van Oyen, M.P., 2004. Agile workforce evaluation: A framework for cross-training and coordination. *IIE Transactions* 36 (10).
- Korhonen, P., Syrjänen, M., 2004. Resource allocation based on efficiency analysis. *Management Science* 50 (8), 1134.
- Martin, C.R.J., Horne, D.A., Chan, W.S., 2001. A perspective on client productivity in business-to-business consulting services. *International Journal of Service Industry Management* 12 (2), 137.
- Napoleon, K., Gaimon, C., 2004. The creation of output and quality in services: A framework to analyze information technology-worker systems. *Production and Operations Management* 13 (3).
- Roels, G., Karmarkar, U., Carr, S., 2010. Contracting for collaborative services. *Management Science* 56 (5), 849–863.
- Strong, R., Chen, T., Zhou, R., (2009). Value measurement in complex service systems. Paper presented at Annual INFORMS Meeting, San Diego, CA.
- Tang, V., (2009). Determining quantitative value in services. Paper presented at Annual INFORMS Meeting, San Diego, CA.
- Van Oyen, M., Gel, G.S., Hopp, W., 2001. Performance opportunity for workforce agility collaborative and noncollaborative work systems. *IIE Transactions* 33, 761–777.
- Winston, W., 1990. *Operations research: Applications and algorithms*. Duxbury Press, Boston.
- Xue, M., Field, J.M., 2008. Service coproduction with information stickiness and incomplete contracts: Implications for consulting services design. *Production and Operations Management* 17 (3), 357–372.